测绘学报
ACTA GEODAETICA et CARTOGRAPHICA SINICA
1999年 第28卷 第3期 Vo1.28 No.3 1999

海洋测量异常数据的检测*

黄谟涛　翟国君　王瑞　欧阳永忠　管铮

摘　要　随着现代科学技术的发展和应用，海洋测量领域已先后推出了多种具有高分辨率和高采集率的新技术测量手段，如用于海底地形测量的多波束测深和机载激光测深系统。作为数据后处理软件系统的一个重要组成部分，我们急需寻求一种有效的异常数据探测方法来对采集到的海量数据进行质量检查和控制。与陆地测量相比较，海洋测量具有明显的动态效应，由于海水阻隔的原因，海洋测量不仅受大气的影响，而且受海水运动和海水物理性质的影响，因此，海洋测量具有比陆地测量更多的噪声干扰源，海洋测量出现粗差的概率也远远大于陆地测量，加上海洋测量进行重复观测比较困难，缺乏必要的几何图形检核条件，这也给解决粗差检测问题增加了一定的难度。
　　本文要解决的核心问题是如何实现在资料检查过程中，自动地发现异常值(包括粗差)的存在，并正确地指出异常值的位置，这就是所谓的异常数据定位问题。以往解决此类问题一般是靠有经验的专业人员，用手工的方式，通过比较数值的大小或分析要素的变化趋势等手段进行判别处理。最近几年，随着计算机技术的普及应用，在海洋测量中，人们相继提出了一些通过计算机自动判别异常数据的方法，这些方法归纳起来可分为两大类，一类是基于经典统计假设理论的误差统计检验法，如文献［1］、［2］、［5］、［6］、［7］和［13］。另一类是基于函数或统计推值的比较判别法，如文献［3］、［4］、［8］、［9］和［15］。应当承认，对上述各种方法加以灵活运用，均能收到一定的效果。但是我们也看到，在实际应用中这些方法仍普遍存在一定程度的不稳定性，其原因是这些方法基本上都是源于经典的最小二乘法，最小二乘估值不具有抗差性，利用这样的估值来构造统计量，并进行统计检验自然很难保证检验结果的可靠性，也就难免出现所谓的异常值“遮蔽”现象[13]，当观测量中存在多个粗差时，这种“遮蔽”现象尤为严重。针对这种情况，本文将尝试应用现代抗差估计理论来提高海洋测量异常数据统计检验的抗差能力，其中心思想是把原来用于检验异常值的推值比较法与抗差估计结合起来，提出一种基于抗差估计的选权迭代插值比较检验法。有关抗差估计的详细理论描述可参阅最近几年发表的相关文献和专著[5，10，14]。
　　由文献［10］得知，对应于经典间接平差模型，基于抗差估计的选权迭代解式为

$$\hat{X}^{(k+1)}=(A^{T}\overline{P}^{(k)}A)^{-1}A^{T}\overline{P}^{(k)}L \qquad (1)$$

式中A为已知系数阵，X为模型待定参数向量，L为自由项，P称为等价权阵。在测量数据处理中，等价权阵一般采用中国学者提出的IGGIII方案[11，14]，即取

$$\overline{p}_i=\begin{cases} p_i & |v_i'|\leqslant k_0 \\ p_i k_0[(k_1-|v_i'|)/(k_1-k_0)]^2/|v_i'| & k_0<|v_i'|\leqslant k_1 \\ 0 & |v_i'|>k_1 \end{cases} \qquad (2)$$

式中$v_i'=v_i/\hat{\sigma}_i$，k_0和k_1可分别取为1.0～1.5和2.0～3.0。

　　本文拟使用简单的加权平均数模型作为判别异常值的推值模型，根据式(1)，可直接写出基于抗差M估计的选权迭代加权平均值计算公式如下：

$$\hat{X}^{(k+1)}=\sum_i \overline{p}_i^{(k)}L_i/\sum_i \overline{p}_i^{(k)} \qquad (3)$$

式中L_i代表被检测点周围邻域内的观测值。式(3)在形式上与传统的加权平均值计算公式完全相同，不同的是这种抗差加权平均值的权在迭代过程中是变化的，它们的值由前一次迭代得到的残差来确定。按式(3)计算平均值时使用的初始权因子p_i，一般可根据观测点到被检测点的距离由下式进行计算：

$$p_i=1/(d_i+\varepsilon)^2 \qquad (4)$$

式中d_i为观测点i至被检测点的水平距离，ε为一小正数，它的作用是防止权函数的分母趋于零。

设由式(3)进行迭代计算达到稳定后的最后结果为$\hat{X}_p$，被检测点上的观测值为L_p(L_p不参加$\hat{X}_p$的计算)，则可求得预测残差为

$$\Delta L_p = \hat{X}_p - L_p \qquad (5)$$

根据ΔL_p的绝对值大小即可以按下式规定的标准对被检测点是否为异常值作出判断：

$$|\Delta L_p| > k\sqrt{\hat{\sigma}_p^2 + m_p^2} \qquad (6)$$

式中$\hat{\sigma}_p$表示推值均方差，m_p为观测中误差，k可以取2或3，视异常值的定义标准高低而定。

利用上述方法进行异常值定位，需要解决的关键问题是推值迭代计算中初始值的选择。本文使用加权中数作为推值模型，属于一维抗差估计问题，因此可以采用观测值中位数作为参数的抗差初值，即可取：

$$\hat{X}^0 = \underset{i}{\mathrm{med}}\{L_i\} \qquad (7)$$

式中med表示取中位数。由式(7)求得的初值具有最高的崩溃污染率(50%)，因此可以保证最后迭代解的抗差性。迭代计算中使用的单位权中误差可由下式计算[12]

$$\hat{\sigma}_0^{(k)} = \underset{i}{\mathrm{med}}\{\sqrt{p_i}\,|v_i^{(k)}|\}/0.6745 \qquad (8)$$

此时有

$$\hat{\sigma}_i^{(k)} = \hat{\sigma}_0^{(k)}/\sqrt{p_i} \qquad (9)$$

在试算过程中我们发现，当由式(2)计算等价权$\bar{P}_i$时，由于初始权p_i(由式(4)计算)之间大小差异较大，离被检测点较近的观测点(这些点对应较大的p_i值，也即较小的$\hat{\sigma}_i$值)，即使是在正常情况下，也可能因为有较大的$|v_i/\hat{\sigma}_i|$值而被拒绝参加计算，结果造成估计效率方面的损失。出现这种现象是由于我们人为选用了比较特殊的初始权p_i，而使用这样的定权公式(见式(4))对提高推值精度又是非常必要的，为了解决两者之间的矛盾，本文将$\hat{\sigma}_i$改由下式直接计算：

$$\hat{\sigma}_i^{(k)} = \underset{i}{\mathrm{med}}\{|v_i^{(k)}|\}/0.6745 \qquad (10)$$

此时初始权p_i对$\hat{\sigma}_i$的确定不再产生直接的影响。试验结果表明，按式(10)计算$\hat{\sigma}_i$从迭代开始到结束都表现得相当稳定，非常适合于抗差迭代解的计算。

为了检验前面介绍的抗差检测异常值方法的有效性，本文首先使用一组模拟数据进行了对比试验，原始数据分布情况及试验结果见表1、表2和表3。在此基础上，本文利用前面提出的异常值检测方法，对我国自行研制的多波束条带测深仪实测数据进行了处理，该批数据分布于38个条带，测点总数共计198 928个，经推值比较共检测出可疑异常值4 742个，占总点数的2.38%。每个条带的测点数以及被检测出的异常值(可疑点)个数和所占百分比的具体情况见表4。模拟数值计算和实际处理多波束测深数据的结果均证实了上述方法的有效性。

关键词　海洋测量　异常数据　粗差检测　抗差估计　可靠性
分类号　P229

The Detection of Abnormal Data in Marine Survey

Huang Motao, Zhai Guojun, Wang Rui
Ouyang Yongzhong, Guan Zheng
(Tianjin Institute of Hydrographic Surveying and Charting, Tianjin, 300061)

Abstract　Blunder detection is a topic of great interest to surveyors and mappers because undetected blunders significantly distort the observed parameters, e.g., soundings. Based on an analysis of the characteristic of marine surveying, a robust method

for the detection of abnormal data (include blunders) in marine measurements is proposed in this paper, which is called the robust interpolation comparison test based on robust M-estimation by an iterative calculation procedure. Some questions involved in the implement of the suggested method are discussed in detail. Compared to the existing methods, the new method has more strong capacity of locating abnormal data. A simulation study and an actual numerical example for the process of multibeam soundings are given to test the performance of the proposed method. The results have illustrated the effectiveness of the method in the detection and identification of multiple blunders. The use of the new method will play an important role in improving the quality and reliability of marine measurements in our country.
Keywords marine survey, abnormal data, blunder detection, robust estimation, reliability

1 Introduction

Compared to terrestrial survey, marine survey is strongly characterized by the dynamic effect. The observations in marine survey are affected not only by atmosphere, but also by the movement and physical property of ocean water. There exist, therefore, more noise sources in marine survey than in terrestrial survey. Taking the shipboard depth sounding for example, the pulse signals emitted from echo sounder could be reflected by floating and swimming living-things (e.g. fishes) and plants during their propagation. These false echoes and additional round trip echoes may result in a big discrepancy between the observed value and the true depth. It is the so-called blundering problem during data acquisition in marine survey. It is obvious that, due to the influences from different kinds of error sources, the blundering possibility in marine survey is much greater than that in terrestrial survey. In addition, it is difficult to make repeat measurement in marine survey at the exactly same position and find out the blunders due to lacking necessary checking-conditions. Thus the blundering problem remains in the data processing.

In the statistical and geodetic literature, however, a blunder is not precisely defined. It is difficult to define since we are not sure if a blunder is a mistake or if the mathematical model is imperfect. Someone calls those observations that are far away from the bulk of the data as blunders or gross errors or outliers. Someone believes that blunders are large in magnitude and much larger than the random errors. Obviously, blunder is a kind of abnormal data (called pseudo-abnormal data). In addition to blunders, there exists another kind of abnormal data in marine survey, i.e., the so-called true abnormal data, which are the true records of the observed parameters. These data are of great value to navigation with safety and design of marine engineering. It has been shown that there exist two kinds of abnormal data in marine survey, and it is essential for us to check and find out these observations in the data processing. However, whether an abnormal observation is a blunder or not should be made a further investigation on the basis of sea state. As to this problem, this paper is not going to make a detail discussion. In this paper, we restrict ourselves to study how to find the existence of abnormal data and then locate them.

The problem of identifying is relatively simple when observations contain a single blunder. But if the observations contain more than one blunder, the problem of identifying becomes more difficult due to the masking or swamping effects. Masking occurs when some abnormal observations go undetected because of the presence of other (usually adjacent) abnormal observations. Swamping occurs when normal observations are incorrectly identified as abnormal ones because of the presence of other (usually remote) abnormal observations. In the past, these abnormal or suspected observations could only be extracted by manually editing blocks of data. With the widespread application of computer for the data processing in marine survey, some methods have been put forward to remove the erroneous data automatically through computer in recent two decades. These methods could be classified into two kinds. One is called statistical test based on some hypotheses, e.g., Luo (1984), Li (1988), Chen (1991), Zhang (1992) and Eeg (1995). The other is called comparison test based on functional interpolation and/or stochastic estimation (collocation), e.g., Herlihy et al. (1992), Ware et al. (1992), Sevilla (1993), Gil et al. (1993) and Zhu (1998). It should be admitted that if the methods mentioned above are reasonably used, part of abnormal observations could be correctly located and removed during the data processing. It is pity that all the methods mentioned above are based on the classical least-squares estimation. And it is well known that the least-squares estimation is not robust, even a single blunder can spoil the solution. If we use these contaminated estimates to construct statistics and then make statistical test, the results will certainly be unreliable. And unavoidably, the problem of masking and swamping effects mentioned above will arise, especially in the multiple-blunder case. To overcome the problem above, this paper makes an attempt to use the robust estimation to increase the reliability of the conventional testing methods for the detection of abnormal data in marine survey. The main idea is to combine the conventional interpolation comparison test with robust estimation, and then a so-called robust interpolation comparison test based on robust M-estimation by an iterative calculation procedure is proposed.

2 The Robust Method

As mentioned above, the conventional least-squares estimation is not robust. It means that the least-squares procedure tends to smooth blunders into good (normal) observations. In other words, the blunders will be spread to other (good and bad) observations and thus the sizes of the actual blunders will be distorted. As a result, incorrect decisions may be derived from the statistical test based on the conventional methods, i.e., a good observation may be rejected or a bad (abnormal) observation may not be detected at all. The advantage of robust estimation is that the negative effect of the blunders on the estimator is greatly softened or even eliminated altogether when making the solution, though the statistical properties of a robust estimator are not clearly defined and the efficiency of the estimator is inferior to a least-squares estimator when no blunders are present.

As we know, robust estimation, on the whole, can be classified into three kinds (see Huang, 1990; Yang, 1993). One is the maximum likelihood type estimation, shortly called M-estimation. Another is the linear combination of order statistics estimation, shortly called L-estimation. And the third one is the rank estimation, shortly called R-estimation. Among the three kinds of robust estimation above, robust M-estimation is the most often used one in geodesy. The method suggested in this paper for the detection of abnormal data is just based on the M-estimation by an iterative calculation procedure. The basic principle of this estimation method with equivalent weights is summarized as follows:

Consider a general error equation

$$V = A\hat{X} - L \qquad (1)$$

$$v_i = a_i\hat{X} - L_i \qquad (2)$$

where $\hat{X}$ is an $m\times 1$ estimate vector of unknown parameters; L is an $n\times 1$ observation vector with an $n\times n$ weight matrix P;V is an $n\times 1$ residual vector of L; v_i and L_i are the ith element of V and L respectively; A is an $n\times m$ design matrix, and a_i is the ith row vector of A. The M-estimation with equivalent weights mentioned above means that a suitable function $\rho(v)$ is chosen to satisfy the following extremum condition (Yang,1993,1994)

$$\sum_i p_i\rho(v_i) = \min \qquad (3)$$

Differentiating the expression above with respect to the unknown $\hat{X}$ yields

$$\sum_i p_i\psi(v_i)a_i = 0 \qquad (4)$$

where $\psi(v_i)$ is the derivative of $\rho(v_i)$. Let

$$\bar{P}_i = p_i\psi(v_i)/v_i \qquad (5)$$

Then by substituting (5) into (4), we have

$$A^T\bar{P}V = 0 \qquad (6)$$

or

$$A^T\bar{P}A\hat{X} - A^T\bar{P}L = 0 \qquad (7)$$

and

$$\hat{X} = (A^T\bar{P}A)^{-1}A^T\bar{P}L \qquad (8)$$

where $\bar{P}$ is called equivalent weight matrix. The calculation of (8) can be made by iterations. Suppose we have obtained the kth estimates of unknown parameters $\hat{X}^{(k)}$ and the residuals $V^{(k)}$, then from equation (8), we get the (k+1)th robust estimates

$$\hat{X}^{(k+1)} = (A^T\bar{P}^{(k)}A)^{-1}A^T\bar{P}^{(k)}L \qquad (9)$$

The robustness of the estimates above is mainly dependent on the determination of the equivalent weights. Some expressions for equivalent weight, which were derived and modified from conventional $\rho(v)$ and $\psi(v)$ functions, have been proposed by statisticians and geodesists (see Huang, 1990; Yang, 1993). Here we use the following equivalent weight function based on the IGG-3 scheme which was originally developed by Zhou (1989) and Yang (1994):

$$\overline{p}_i=\begin{cases}p_i & |v_i'|\leqslant k_0\\ p_i k_0[(k_1-|v_i'|)/(k_1-k_0)]^2/|v_i'| & k_0<|v_i'|\leqslant k_1\\ 0 & |v_i'|>k_1\end{cases}\qquad(10)$$

where $v_i'=v_i/\hat{\sigma}_i$, and $\hat{\sigma}_i^2=\hat{\sigma}_0^2/p_i$; $\hat{\sigma}_0^2$ is the estimate of unit weight variance. The constant k_0 is proposed to be 1.0～1.5 and k_1 to be 2.0～3.0.

The so-called robust interpolation comparison test suggested in this paper can be described as a three-step process. First each raw observation takes a turn being the comparison observation, or the observation currently under evaluation. Then the comparison observation is examined relative to the robust weighted average of the neighbor observations taken from the neighborhood of the comparison point. And finally, a criterion is used to accept or reject the observation based on this comparison.

Our problem here belongs in one dimension robust estimate. According to equation (9), the expression of the robust weighted average by an iterative calculation procedure can be directly written as

$$\hat{X}^{(k+1)}=\sum_i \overline{p}_i^{(k)}L_i/\sum_i \overline{p}_i^{(k)}\qquad(11)$$

where $L_i(i=1,2,\cdots,\ n)$ represents the neighbor observations taken from the neighborhood of the comparison observation L_p. It can be seen from (11) that, formally, the general formula of the robust weighted average is identical to that of the conventional weighted average. The only difference is that the weight factor p_i of the conventional weighted average is now replaced by the equivalent weight factor $\overline{P}_i$ of the robust weighted average. Whereas it is this substitution that can make the robust weighted average resist the influence of blunders.

As usual, the initial weights $p_i(i=1,2,\cdots,\ n)$ used for the calculation of (10) can be assigned according to their horizontal distances to the interpolated point as follows:

$$p_i=1/(d_i+\varepsilon)^2\qquad(12)$$

where d_i indicates the horizontal distance between the observation point L_i and the interpolated point (i.e. comparison point) L_p. ε is an arbitrary constant to deal with p_i approaching infinitude as the denominator of the weight function approaches zero. In practical computation, we can set $\varepsilon=0.01$. It can be seen from (12) and (10) that the observations near the interpolated point, in normal case, have a great influence on the interpolation. Whereas when some observation is contaminated by blunder, the residual of the observation will increase, and the corresponding equivalent weight calculated from (10) will decrease. It means that the influence of the abnormal servation on the interpolation will be descending even if the observed point is so much close to the interpolated point. When $|v_i|>k_1\hat{\sigma}_i$, $\overline{P}_i=0$, that is to say, the significantly abnormal observation has no influence on the interpolation.

Suppose $\hat{X}_p$ to be the convergence value of equation (11) and L_p the corresponding comparison observation (L_p does not take part in the calculation of $\hat{X}_p$). Then the predicted residual can be defined as (note that $\hat{X}_p$ here is one dimension parameter)

$$\Delta L_p=\hat{X}_p-L_p\qquad(13)$$

Finally, the absolute value of ΔL_p can help us make a decision concerning the observation L_p based on a comparison between $|\Delta L_p|$ and a critical value or threshold ΔL_{max}. Here the critical value ΔL_{max} is defined to be the maximum acceptable residual. The magnitude of ΔL_{max} is dependent on the accuracy of observation and the perfection of interpolation model. In practical application, ΔL_{max} is usually chosen as two or three times the standard deviation of observation.

The key to the above robust method for the detection of abnormal data lies in the determination of starting values of the unknown parameters. For some cases, the starting solution even can determine whether a usable M-estimator is obtained or not. In our opinion, the starting value of robust solution should be chosen to be robust in order to get a reliable convergence. The median of observations, in this paper, is suggested to be taken as the starting value of $\hat{X}$ for the calculation of (11). That is

$$\hat{X}^0 = \underset{i}{\mathrm{med}}\{L_i\} \tag{14}$$

where medi denotes median over i. It is known that the estimator defined by equation (14) has the highest possible breakdown point of 0.5 (Yang, 1993). So it can ensure the stability of the iterative procedure. Yang (1997) suggested that the variance factor used for the iterative procedure be calculated by

$$\hat{\sigma}_0^{(k)} = \underset{i}{\mathrm{med}}\{\sqrt{p_i}\,|v_i^{(k)}|\}/0.6745 \tag{15}$$

and

$$\hat{\sigma}_i^{(k)} = \hat{\sigma}_0^{(k)}/\sqrt{p_i} \tag{16}$$

According to our experiences in application, it is found that when the equivalent weights are calculated through equation (10), the observations near the comparison point might be rejected to take part in the calculation of the interpolation due to their having a larger weight factor p_i, i.e., a smaller variance factor $\hat{\sigma}_i$, and a bigger ratio of $|v_i|/\hat{\sigma}_i$. As a result, it may cause the loss of interpolation efficiency. The reason for the above result is that we have chosen a special initial weight function (see equation (12)) which gives a great difference among the observations. Whereas such an initial weight function is essential to the improvement of interpolation accuracy in the normal case. In order to resolve the contradiction above, here the variance factor, by taking the observations as parameters with equal accuracy, is suggested to be calculated directly by

$$\hat{\sigma}_i^{(k)} = \underset{i}{\mathrm{med}}\{|v_i^{(k)}|\}/0.6745 \tag{17}$$

In this case, the initial weight factor p_i has no more direct influence on the determination of $\hat{\sigma}^{(k)}_i$. And as a result, all of the variance factors $\hat{\sigma}_i$ are kept stable in the whole iterative procedure.

3 Numerical Examples

In order to show the efficiency of the robust method proposed above, two numerical examples are given to test its performance. The first example is a simulation study where a set of data are simulated, which consist of twenty-five observations with an equivalent interval and follow a normal distribution $N(\mu=10, \sigma^2=1)$. The order numbers and magnitudes of the simulated observations are given in Table 1.

表1

Tab.1 The distribution of the simulated data (order number/magnitude)

Order number	Magnitude	Order number	Magnitude	Order number	Magnitude	Order number	Magnitude	Order number	Magnitude
1	8.18	6	8.49	11	9.19	16	10.20	21	10.47
2	8.42	7	8.64	12	9.38	17	10.55	22	10.81
3	9.00	8	9.14	13	9.93	18	10.76	23	11.11
4	9.49	9	9.79	14	10.46	19	11.11	24	11.69
5	10.16	10	10.54	15	10.96	20	11.49	25	11.94

Then each of the initial observations takes a turn being the interpolated point. And two interpolations for each point are performed by using the robust weighted average and the conventional weighted average, respectively. Based on the predicted residuals (see equation (13)), two standard deviations corresponding to the two methods are computed as: σ (robust)=0.50 and σ (conventional)=0.53. It is shown that, in the normal case (without blunders), the prediction accuracies of the two methods are nearly the same. To show the feasibility of the recommended procedure, four blunders are added to four initial observations, respectively. Two cases are treated:

Case one-4σ size of blunder is added on observation 8, 12, 14 and 18 at the same time.

Case two-4σ size on observation 8 and 10σ size on observation 12, 14 and 18, respectively.

The predicted residuals corresponding to the two cases above by using the interpolation methods of robust and conventional weighted average are calculated and given in Tab.2 and Tab.3, respectively.

表2

Tab.2 The predicted residual errors corresponding to case one

Robust	Conventional	Robust	Conventional	Robust	Conventional	Robust	Conventional	Robust	Conventional
-0.95	-1.44	-0.82	-1.40	-0.73	-1.36	-0.42	-0.92	0.12	-0.30
-0.77	-1.39	-0.84	-1.89	3.54	2.91	-0.64	-1.36	0.23	-0.26
-0.68	-1.39	3.59	3.03	-0.69	-1.85	3.96	3.57	0.27	-0.46
0.03	-0.44	0.10	-0.78	4.14	3.76	0.02	-0.74	0.43	0.05
0.04	-0.30	0.26	-0.25	0.25	-0.47	0.43	0.12	0.75	0.56

表3

Tab.3 The predicted residual errors corresponding to case two

Robust	Conventional	Robust	Conventional	Robust	Conventional	Robust	Conventional	Robust	Conventional
-0.94	-1.90	-0.80	-1.86	-0.68	-1.83	-0.42	-1.38	0.13	-0.76
-0.73	-2.08	-0.75	-2.91	3.56	2.05	-0.58	-2.37	0.26	-0.96
-0.52	-2.59	9.59	8.51	-0.51	-3.82	9.98	9.05	0.31	-1.66
0.13	-1.26	0.27	-2.37	10.19	9.07	0.21	-2.34	0.51	-0.77
0.10	-0.92	0.30	-1.09	0.32	-1.75	0.60	-0.73	0.90	-0.06

As shown in Tab.2 and Tab.3, in both cases, the comparison tests based on the robust interpolation method proposed in this paper can always find the blunders correctly, even for small sizes (4σ). Conversely, the tests based on the conventional interpolation are not so satisfactory, especially in case two, where the predicted residual of good observation (number 13) is even larger than that of bad observation (number 8). It is shown that when there exist multiple blunders and the sizes of them are not uniform, the problem of masking and swamping effect mentioned in the previous section may arise by using the conventional method, whereas it does not by using the proposed robust method here.

The second example is a set of actual soundings from the Chinese-developed H/HCS-017 swath bathymetry system. The data set consists of 38 swathes and a total of 198,928 soundings. By using the robust interpolation comparison test to check the data set, a total of 4 742 soundings are found to be abnormal, which account for a 2.38% of the total soundings. The statistical results of test for the data subsets of each swath are given in Tab.4.

表4

Tab.4 The statistical results of detecting abnormal observations from multibeam sounding

Swath number	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
Total soundings	4688	4384	4144	6272	5200	4720	4448	4560	4768	4720	4272	3904	3984	5472	4576	4496	4128	4544	4224
Abnormal soundings	191	195	56	243	135	114	134	85	127	137	17	47	64	18	37	75	113	127	68
Percentage	4.06	4.45	1.35	3.73	2.60	2.42	3.01	1.86	2.66	2.90	0.40	1.20	1.61	0.33	0.81	1.67	2.74	2.79	1.61
Swath number	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38
Total soundings	4144	4256	4176	6048	7664	7872	6224	5024	3952	5808	7376	6272	4208	7408	6368	5728	7744	6992	4160
Abnormal soundings	78	102	35	273	256	252	230	181	87	68	265	181	0	199	109	67	159	153	73

Percentage /(%)	1.88	2.40	0.84	4.51	3.34	3.20	3.70	3.60	2.20	1.17	3.59	2.89	0	2.69	1.71	1.17	2.05	2.19	1.75

As shown in Tab.4, the percentages of abnormal observations found in this example are a little larger than those listed in Herlihy (1992). Having made a further analysis on the detected abnormal data, we find that more than 90% of abnormal observations are located in the border areas of each swath. It gives an indication of where improvement should be made in this new sounding system in order to increase the accuracy and reliability of soundings.

4 Conclusions

High volume data acquisition techniques for mapping the seabed, e.g., multibeam echosounding system and airborne laser depth sounding system, have recently become available and adopted for use in China. These systems have a number of features in common. A high data rate is one of them. As an important part of the quality control of data, it is essential for us to develop a valid method in time for detecting automatically the abnormal data from the high volume observations. This paper has introduced the theory of robust estimation to the data processing of marine survey. The purpose is to promote people to pay more attentions on the quality control and reliability of data in marine survey in China. Our proposed procedure for the detection of abnormal data is a composite of the robust estimation and the conventional interpolation comparison. The simulation study and the practical process of swath bathymetry data have proven the new method to be an effective procedure of detecting abnormal data. It is especially impressive for the multiple blunder case.

作者简介：黄谟涛，男，38岁，高级工程师，硕士，现从事海洋重力研究。
作者单位：黄谟涛　翟国君　王瑞　欧阳永忠　管铮(天津海洋测绘研究所，天津，300061)

References

1　Chen S J and Ma J R. The Application and Analysis of Marine Observation Data. Beijing: Marine Publishing House, 1991 (in Chinese)
2　Eeg J. On the Identification of Spikes in Soundings. International Hydrographic Review, 1995, LXXII(1):33～ 41
3　Gil A J, et al. A Method for Gross-error Detection in Gravity Data. International Geoid Service, Bulletin, 1992(2): 25～31
4　Herlihy D R, et al. Filtering Erroneous Soundings from Multibeam Survey Data. International Hydrographic Review, 1992, LXIX(2): 67～76
5　Huang Y C. Robust Estimation and Data Snooping. Beijing: Publishing House of Surveying and Mapping, 1990 (in Chinese)
6　Li D R. Error Processing and Reliability Theory. Beijing: Publishing House of Surveying and Mapping, 1988 (in Chinese)
7　Luo N X. Data Processing and Survey Error. Beijing: Measurement Publishing House, 1984 (in Chinese)
8　Sevilla M J. Analysis and Validation of the D.M.A Gravimetric Data of the Mediterranean Sea. GEOMED Report-3, 1993. 78～101
9　Ware C, et al. A System for Cleaning High Volume Bathymetry. International Hydrographic Review, 1992, LXIX(2): 77～94
10　Yang Y X. The Theory and Application of Robust Estimation. Beijing: Bayi Publishing House, 1993 (in Chinese)
11　Yang Y X. Robust Estimation for Dependent Observation. Manuscr Geod, 1994(19):10～17
12　Yang Y X. Estimators of Covariance Matrix at Robust Estimation Based on Influence Functions. ZfV, 1997(4): 166～174
13　Zhang F R and J T Zhang. The Distribution and Statistical Test of Survey Errors. Beijing: Measurement Publishing House, 1992 (in Chinese)
14　Zhou J W. Classical Theory of Errors and Robust Estimation. Acta Geodaetica et Cartographic Sinica, 1989(2): 115～120
15　Zhu Q and D R Li. Error Analysis and Processing in Multibeam Soundings. Journal of Wuhan Technical University of Surveying and Mapping, 1998(1):1～ 4 (in Chinese).

收稿日期：1999-02-25，截稿日期：1999-04-30